Roger H. Reeves, Ph.D.
Department of Physiology
Johns Hopkins University
School of Medicine
rreeves@welchlink.welch.jhu.edu

Genetic, Physical and Biological Mapping in Mouse

Cartography is the science or art of making maps. There are many approaches to building genetic maps and these different approaches provide diverse, complementary types of information that have different applications to specific scientific or medical questions. In genetic maps, as in any type of map, the competing elements are scope and resolution; a procedure providing rapid coverage of the entire genome will necessarily be low resolution, while the DNA sequence of a single gene represents a tiny percentage of the whole genome. Complementary mapping methods allow production of increasingly detailed frameworks that permit assembly of more and more highly resolved maps. Currently, the International Genetics community has undertaken the Human Genome Initiative, whose ultimate goal is to determine the sequence of the entire human genome and of the genomes of selected model organisms. In this lecture we will introduce the mouse as a model organism for genetic analysis to establish the theoretical underpinnings of recombinational maps, and then consider practical aspects of mapping genes in mice and using this information to better understand genetic organization in humans.

Today's lecture will address two questions:

1) What can be learned from mouse genetic, physical and sequence mapping that won't be learned from *Homo sapiens*?

2) How do technical approaches to mapping in mouse contrast with those used for human beings, and how do they allow for different types of information to be obtained?

Why map in mouse?

1. Pragmatic: Mouse is a selected model organism for the human genome project, it's therefore necessary to generate sequence-ready template for sequencing. Since non-map-based approaches (e.g., BAC end sequencing) on a genomic scale have been largely eschewed at this point, that leaves map-based approaches by default. In addition, maps already exist that contain substantial information for clone ordering.

2. Comparative mapping. Genes be discovered in mouse and mapped far more efficiently than in outbred mammalian species like *Homo sapiens*, and this genetic information can be related directly to humans through comparative mapping.

3. Basic clues to central questions in genetics will be addressed, if not solved altogether, by comparing the genome sequences of different mammals. Included in this category would be investigations into the mechanisms by which genomes evolve. A fundamental question remains as to the function of the 80% of the mammalian genome that which is unique sequence DNA but does not encode genes (of the types we're familiar with) or major structural elements (centromeres, telomeres, etc.).

4. Gene function (a.k.a. Functional Genomics - originally known in two earlier manifestations as "Genetics" or as "Physiology"). The mouse is a living encyclopedia of known gene functions and a repository for the (mostly) unknown gene functions that produce a developing, metabolizing, reproducing mammal. Both targeted and genome-wide mutagenesis strategies now being employed represent a new kind of mapping directed toward the elucidation of gene and genome function by the time-honored genetic approach of comparing mutant and normal phenotypes.

How is mouse mapping different from mapping in humans?

Mice provide many advantages as a model system for mammalian genetics that stem from short generation time, large litters, relatively low cost of care, and the ability to create **inbred strains of mice** where every individual (of the same sex) is genetically identical. The existence of inbred strains enhances the information content of genetic maps and allows several types of mapping to be accomplished that are not readily accomplished or are impossible in other species.

Recombinational mapping is enhanced by the ability to do phase known crosses with complete heterozygosity of every individual at every locus. This allows the production of a recombinational map that is not based in statistics but is actually based directly on the physical sites at which recombination occurs. Excluding typing errors (which are minimized when only two specific outcomes are possible for each typing reaction), the mouse recombinational map produces an unequivocal order for any loci separated by recombination, and thus provides a scaffold for assembly of physical maps that is much more highly resolved than the maps for humans. The ability to examine the outcome of the same meiosis over and over again (to thousands of times) provides a means of addressing individual variability in maps that is unique to the mouse. Finally, both defined backcrosses and intercrosses, and specialized mapping "reagents" like recombinant inbred (RI) strains, congenic strains and recombinant congenics provide signficant shortcuts to the identification of quantitative trait loci (QTLs) and modifier genes, specific alleles of which are carried on defined (inbred) backgrounds.

THEORETICAL ASPECTS OF MAPPING

1. Types of maps

Table 1. Types of maps

- Cytological
  - In situ
  - Somatic cell hybrids (translocation hybrids)
- Recombinational (statistical)
  - Pedigrees analysis, MLE (cumulative info - CEPH)
  - Backcross/intercross (model systems)
- Physical
  - Radiation hybrids
  - Chromosome-specific libraries
  - Linking/jumping
  - Contigs (clones: YACs, P1,)
  - PFGE
  - Sequence
- Comparative mapping

Magnitude of the problem.

The mouse and human genomes are about 3 X 10^9 bp in length (haploid; mammals range from ~1.5 - 5.5 X 10^9 bp per haploid genome). The human genome measures ca. 3000 cM, so 1 cM in humans is approximately 1Mb; the mouse is about 1500 cM, so 1 cM in mouse is ~2 Mb. Mammalian genomes probably encode 50,000-100,000 genes, or one every 30-60 kb.

2. Types of markers

Mapping involves assigning a distinct (distinguishable) marker to some fraction of the total genome. Thus, mapping can be considered from the perspective of the types of markers used or the procedures used to define sub-sets of the genome.

Table 2. Types of markers .

- DNA probe
  - Unique
  - Mid-repetitive
  - VNTR
- Simple sequence repeat
- Isozyme
- Antibody
- Biochemical assay
- Visible phenotype

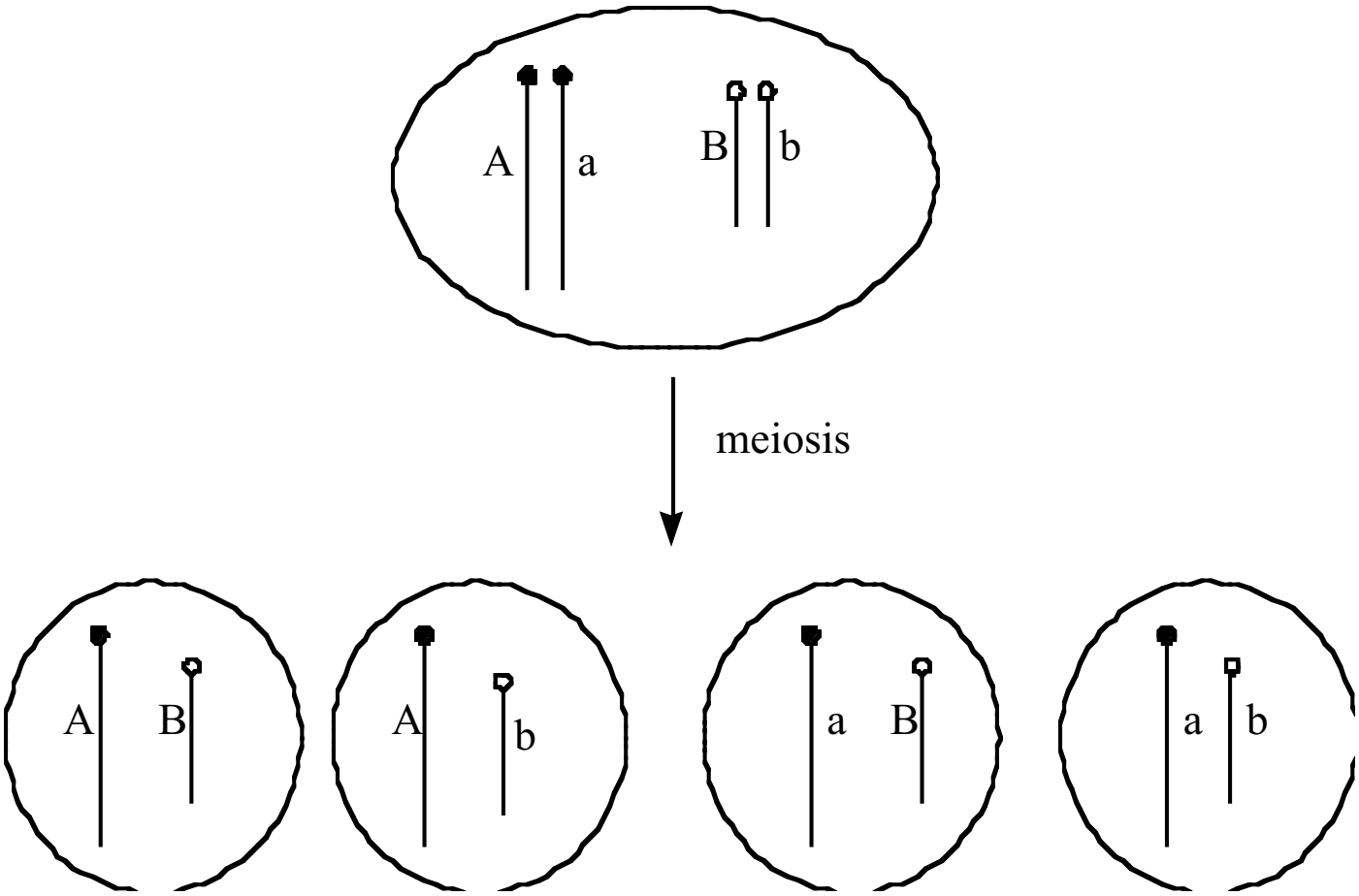
POLYMORPHISM: In order for two markers to be followed in a cross, they must be distinguishable from each other, i.e., they are *polymorphic* or *variant.*

3. Linkage: are two different genes (markers) inherited together or independently?

Genetic information is encoded in DNA and packaged into chromosomes. Eukaryotic organisms are diploid at some stage in their life cycles, that is, they have two copies of each chromosome (homologs). The two copies of a given chromosome have the same overall organization, that is they contain the same subset of the organism's genes

in the same order. However, the two chromosomes are usually not identical, because the same gene may have different variants in the population. This variation allows the genes or markers on the pair of homologous chromosomes to be followed through multiple generations.

A given pair of markers may be *linked* (inherited together) or *unlinked* (inherited independently). The following example shows the results of *independent assortment* (random segregation) of unlinked markers in (A) and the expected pattern of non-random segregation or co-segregation of *linked* markers in (B).



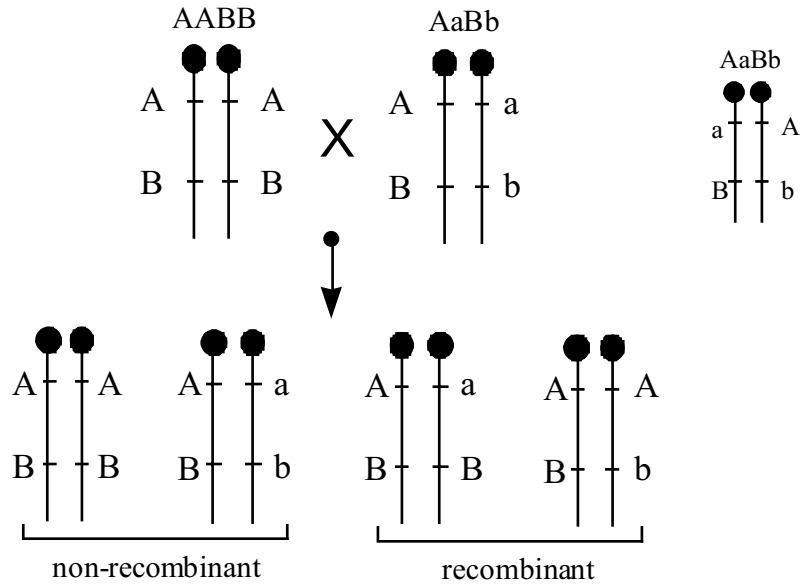
4. Recombinational maps are based on meiotic recombination of polymorphic alleles/markers in a pedigree. Statistical methods are used to determine the frequency of recombination to determine the distance between genetic markers.

Recombinational maps are used to determine the relative order and *genetic distance* between markers. Genetic distance is a statistical measure of the frequency of recombination between two markers measured in units called centiMorgans (cM; 1 cM = 1% recombination between two markers). *Recombination frequency* is a biological property affected not only by the distance between two markers, but also by biological factors such as the age, sex and genetic background of the individual in which the observed meiosis occurs. Model systems provide the ability to measure genetic distance by the direct method. Mapping in populations of random breeding, outbred species is accomplished using a much more complex procedure involving analysis of pedigrees to determine a maximum likelihood estimate (MLE) of relative gene position and distance. The key information required for recombinational mapping is a knowledge of **phase** and **polymorphism** for the markers being typed.

DIRECT METHOD. This method is used where a cross has been established such that markers of interest are polymorphic and phase is already known.

a. Two point test cross determines genetic distance between two genes as **recombination fraction (3)**, which is simply the number of recombinants divided by the number of meiotic events observed. Test crosses rely on the ability to identify unequivocally all recombinants and all non-recombinants. This condition is met in a **phase known double backcross**: one parent is doubly homozygous, one is doubly heterozygous and the arrangement of the four alleles (phase) is known.

		AB
(s)	AB	AABB
	ab	AaBb
(r)	Ab	AABb
	aB	AaBB



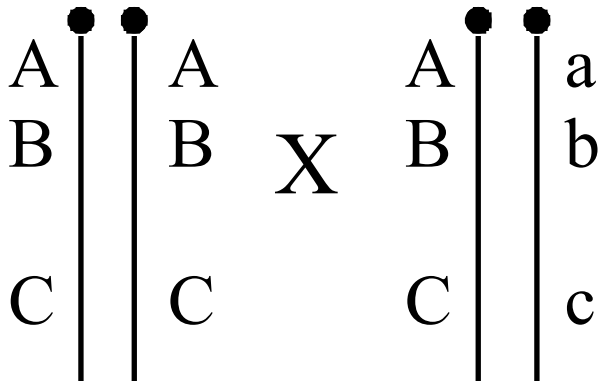
b. Backcrosses segregating multiple markers allow determination of distance and order. The mathematical probability of double crossovers is the product of the probability for recombination at $(3)_{a\text{-}b}$ and $(3)_{b\text{-}c}$, so order is determined by minimizing the number of double crossovers. The actual frequency of double crossovers will be less then predicted by this formula: this biological property is known as **interference**.

A	B	C			Genotype
M	M	M	27	55	AABBCC
T	T	T	28		AaBbCc
M	T	T	4	9	AABbCc
T	M	M	5		AaBBCC
M	M	T	16	35	AABBCc
T	T	M	19		AaBbCC
M	T	M	1	1	AABbCC
T	M	T	0		AaBBCc
			100		

M = homozygous
T = heterozygous

A to B = 10/100 = 0.10
B to C = 36/100 = 0.36
A to C = 46/100 = 0.46

Expected frequency of dbl. x-overs: 0.1 X 0.36 = .036



Resolution of a genetic map is determined by the number of informative meiotic events typed. High levels of allelic variation are achieved by the use of different species or subspecies of *Mus*. In theory, an interspecific backcross could be designed to provide twice the information of a "classic" backcross between inbred strains because it will provide a three allele system. This has not been realized in practice, partly due to the fact that many interspecific crosses result in sterile offspring. This same principle can be applied to construction of recombinational maps in outbred species where a fertile F_1 hybrid can be produced (e.g., cats, cattle). This approach can be used to map outbred species for which 1) fertile hybrids can be formed with another species or subspecies and 2) defined genetic crosses can be established.

> I am a single well-groomed Indian gaur interested in whole grain foods and interspecific hybridization. You are a like-minded *Bos taurus*, less than 2200 lbs., preferably into ag. research. Holsteins preferred. Box 234 MOOO.

Since markers are typed relative to other markers, map information for a given individual is cumulative. This has led to the use of common sources of DNA from defined pedigrees that are screened by many different laboratories (e.g., "the CEPH families" and "the Venezuelan pedigrees" for humans; "Neal and Nancy," "EUCIB" and "JAX BSS" for mouse).

RECOMBINATIONAL MAPPING

Marker type: Virtually any variable trait, including DNA (any type), isozyme, biochemical, cytological marker, antibody, morphological marker, disease. For humans, many markers are uninformative, i.e., they don't vary, so highly variable sequences throughout the genome are now a preferred marker type.

Localization: To the distance between the closest proximal and distal markers that recombine. A 5-year goal of the human genome initiative is a map with an average resolution of 2-3 cM.

Resolution: Resolution is determined by the number (and distribution) of linked polymorphic markers and the number of informative meioses carried in available members of the pedigree. For humans, reaching a map with an average of one marker per 3 cM (~3 megabase) is being undertaken as a major international "collaboration" consuming hundreds of millions of research dollars per year. For experimental organisms like the mouse, single laboratories map regions of chromosomes to 0.2 cM resolution.

PEDIGREE ANALYSIS. In outbred species, such as humans, maps are based on statistical analysis of linkage. These maps provide a probabilistic estimate of order and distance of markers. Human recombinational maps are primarily constructed from analysis of pedigrees using a *maximum likelihood estimate* (MLE) to ask, for two genes A and B, what is the probability that the combination of genotypes seen in this set of progeny would occur if A and B are separated by some distance (3)? An iterative solution is then applied using different values of 3 to find the value with the highest probability.

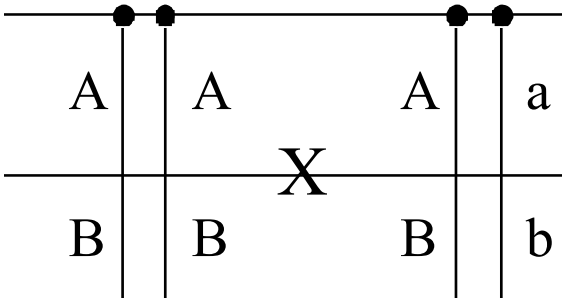
The goal of linkage analysis is to establish whether 3 is < 0.5; if so genes are linked. The expectation of this event can be stated as a ratio of the likelihood that two genes are linked at a certain distance to the likelihood that they are unlinked, which is called the **odds ratio**. For convenience in computing, the decimal logarithm of the odds ratio, Z(3), is routinely used in calculations. Z(3) is the known as the **lod score**.

$$\text{odds ratio} = \left[L(3)/L(0.5) \right]$$

$$\text{lod score} = Z(3) = \log [L(3)/L(0.5)]$$

This equation is used to determine the likelihood that an observed collection of genotypes would occur among a particular set of offspring if the genes being analyzed were separated by a distance equal to some optimal value of 3.

Simple example: AABB X AaBb ->



2 AABB non-r (s)
1 AaBb non-r (s)
1 AABb r
0 AaBB r

What is the 3 between A and B?

i. Direct method: 3 = 1/4 = 0.25
ii. MLE: What is the likelihood that we would obtain 3 non-recombinants and 1 recombinant if A and B are separated by 3? [The likelihood depends on whether they're linked, how close they are, and on how many offspring were typed.] For this e.g., one recombinant occurs with the likelihood 3, and 3 non-recombinants occur each with the likelihood of (1-3). Substituting into the equation above,

$$Z(3) = \log [3(1-3)^3/(0.5)^4]$$

To solve this equation, substitute different values of 3 from 0 to 0.5. The most likely value for 3 is that where Z(3) is largest. For the simple calculation, the highest lod score is obtained when 3 = 0.25 (Table 1). To analyze actual data from genotypes, both alleles have to be considered and the equation is modified:

$$Z(3) = r[\log(23)] + s(\log[2(1-3)])$$

The answer for the simple example is the same (Table 2). If the markers are unlinked, Z(3) is largest at 3 = 0.5 (Table 3). The magnitude of Z(3) provides a measure of confidence in the result. Generally, a lod score of >3.0 (odds of 1000:1) is considered a threshhold for linkage, while a maximum Z(3) < -2.0 is considered proof of non-linkage. Confidence is affected by sample size.

Simple calculations can be done by hand, but actual experimental data of any magnitude is always analyzed using computer programs such as LIPED, LINKAGE and MAPMAKER, and more recently, LINKMAP and FASTMAP. Some of these programs permit analysis of more than two loci, and are thus used for multipoint linkage analysis to determine the most likely order for 3 or more genes. Multipoint analysis rapidly becomes "computation intensive" as more loci are added. (For 5 loci there are 10 values of 3 [3 for the distances a to b, a to c, a to d...d to e] and 60 possible symmetrical orders).

Table 2. Sample LOD calculations.

Lod score, simple example for 1 recombinant and 3 non-recombinants

$Z(\theta) = \log[\theta (1- \theta)^3 / 0.0625$

for s=3, r=1

Theta	Z (θ)
0.01	-0.809
0.05	-0.164
0.1	0.067
0.2	0.214
0.25	0.227
0.3	0.217
0.4	0.141
0.5	0.000

LOD scores for phase known markers, 1 recombinant and 3 non-recombinants

$Z(\theta) = r*\log(2\theta) + s*\log[2(1-\theta)]$

For s=75, r=25

Theta	Z (θ)
0.01	-20.224
0.05	-4.093
0.1	1.671
0.2	5.360
0.25	5.681
0.3	5.413
0.4	3.516
0.5	0.000

Lod scores for phaseknown markers, 1 recombinant and 3 non-recombinants

$Z(\theta) = r*\log(2\theta) + s*\log[2(1-\theta)]$

for s=3, r=1

Theta	Z (θ)
0.01	-0.809
0.05	-0.164
0.1	0.067
0.2	0.214
0.25	0.227
0.3	0.217
0.4	0.141
0.5	0.000

2 recombinant and 3 non-recombinants

$Z(\theta) = r*\log(2\theta) + s*\log[2(1-\theta)]$

for s=2, r=2

Theta	Z (θ)
0.01	-70.115
0.05	-36.062
0.1	-22.185
0.2	-9.691
0.25	-6.247
0.3	-3.786
0.4	-0.886
0.5	0.000

In the example presented above, we assumed that phase was known, but in fact, it could not be determined from analysis of only the individuals presented. Phase can sometimes be determined by typing the grandparents. For example, if the parents of the double heterozygote were AABB and aabb, then the phase was as shown. Large pedigrees used for constructing a human linkage map include such three generation “nuclear families”. Phase can sometimes be deduced, as well; if 100 offspring show an apparent 3 of 0.1 assuming a certain phase, then they would by definition demonstrate 90% recombination assuming the opposite phase. This is statistically unlikely, (unlinked genes will show 50% recombination on average) so phase can be assumed. However, there are biological situations that could result in this type of skewing, so there is a cost in uncertainty associated with such assumptions.

Mapping human genetic diseases

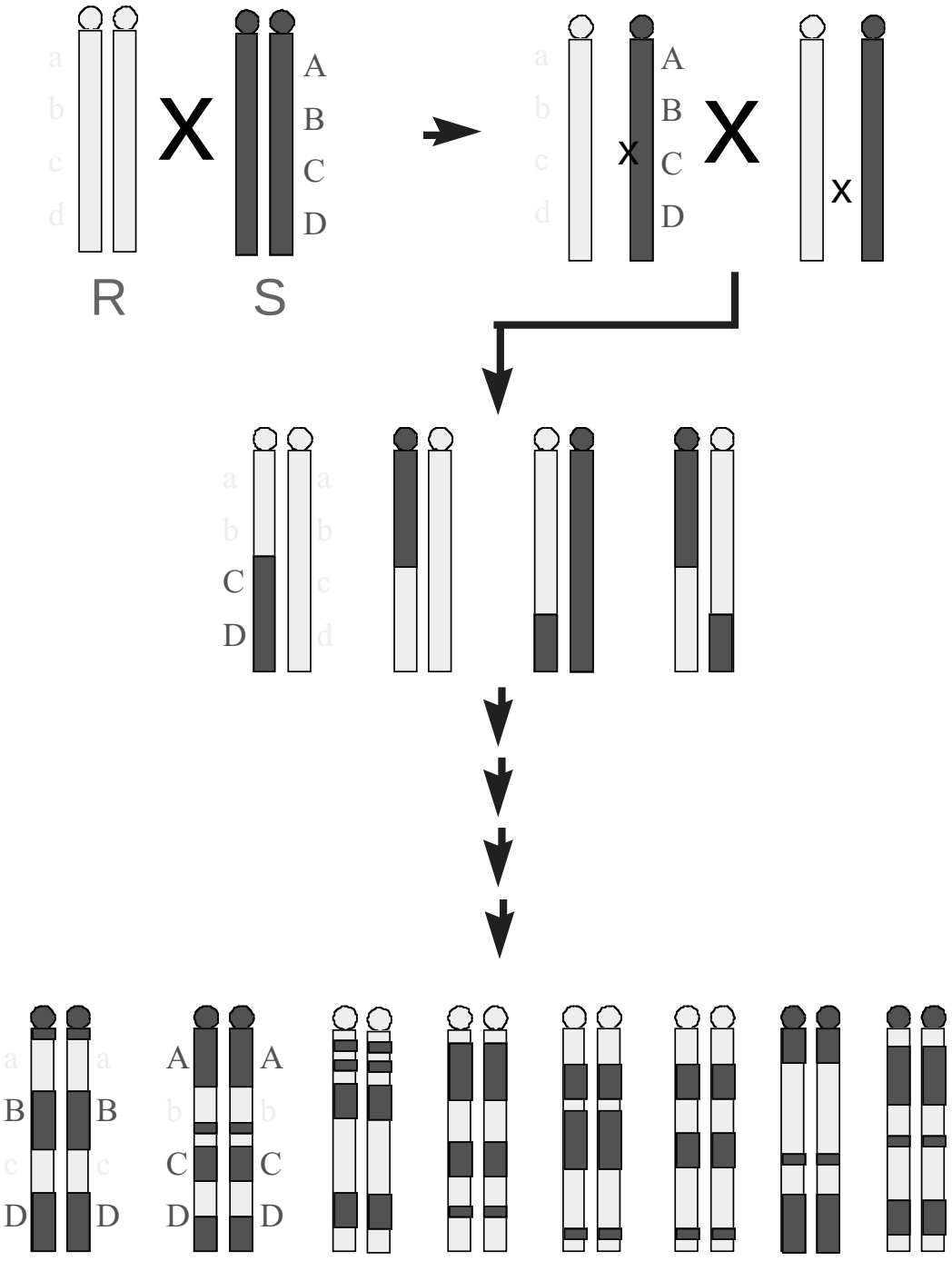
Genetic diseases are mapped by following segregation of the disease phenotype. Factors affecting the course of the disease can often be taken into account by linkage analysis computer programs, but must be intuited correctly to avoid erroneous conclusions. The following conditions must be met:

1. The mode of inheritance must be known.
2. All phenotypic variation is assumed to be due to a single gene and random environmental influences.
3. There are no epistatic interactions (e.g., having the disease doesn't cause a disproportionate number of + type or mutant gametes to be transmitted).
4. Penetrance (age of onset).
5. Allele frequency (PIC) - differs in different populations.
6. Small pedigrees, low polymorphism, paucity of markers near the disease gene.
7. Sex - recombination is higher overall in females, but regions of elevated male recombination occur. By extrapolation from studies with inbred mice, differences also occur between families. These are likely to be small, but may be critical in genetic counseling.
8. Disease heterogeneity/ full and correct understanding of the etiology and pathology of the disease.
9. 100% accurate diagnosis.
10. Rate of interference.

Special case: mapping quantitative trait loci (QTLs).

One of the most relevant problems to human disease gene mapping is the occurrence of diseases that are caused or heavily influenced by multiple genes located at many different sites around the genome. Progress in the development of methods to study inheritance of complex genetic traits has come through the use of model systems. Highly informative markers distributed throughout the genomes of rat and mouse have been useful in identifying inheritance of hypertension, diabetes and obesity. In these applications, backcrosses are performed using strains that vary in expression of the trait. Progeny are screened for the degree of expression and then for inheritance of markers spanning the genome. Where a marker falls near a gene that has a strong influence on the trait, there will be an indication of linkage to the disease in an individual expressing that trait. This type of analysis is especially important in human beings, which are highly outbred. The genetic variability among individuals not only determines inheritance or not of genetic disease, but also clearly affects the general health and response to specific pathological situations occurring in that individual.

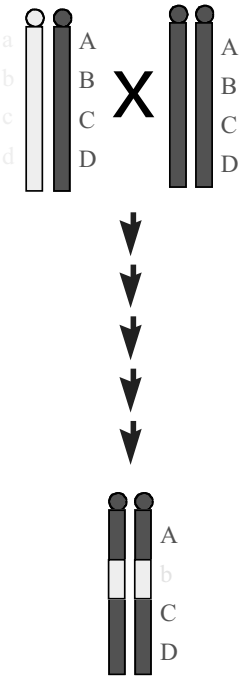
5. Recombinant inbred (RI) mouse strains are produced by inbreeding for >20 generations.
RI strains are useful for rapid localization (mapping) of markers based on strain distribution patterns (SDPs). RI strains can be used to map the genes responsible for phenotypes that are visible only in animals (e.g., metabolic differences, predisposition to disease, etc.), including recessive and, to some degree, polygenic traits.



R S R S R R R R

Congenic strains transfer single genes to different genetic backgrounds. They have been used extensively in immunogenetic studies. A variety of mutant stocks are available that are maintained as congenics after breeding to homozygosity, or that arise spontaneously on inbred backgrounds resulting in **coisogenic** strains that differ only at the site of the mutation.

This process can be carried out beginning with an RI strain that shows a particular phenotype of interest, crossing to a genetic background on which the response is modified to from recombinant congenic mice. This has been instrumental in mapping genes contributing to tumor progression in mouse models, such as *Min1* gene predisposing to multiple intestinal neoplasia, now known to be caused by a mutation in the murine homolog of the adenomatous polyposis coli (*APC*) gene.



Chromosome: 7

AXB	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Fau-ps3	B	B	A	B	A	A	A	B	B	B	A	A	A	A	A	.	A	B	B	B	B	.	B	B	.
Iapls1-38	A	B	A	B	B	A	A	B	B	B	B	A	A	A	A	.	B	B	B	B	B	.	A	B	.
Iapls1-46	B	A	A	A	A	A	B	B	A	B	A	A	A	A	A	.	A	B	B	B	B	.	B	B	.
D7J3	B	A	A	B	A	A	B	B	A	B	B	A	A	A	B	.	B	B	B	B	B	.	B	B	.
Iapls3-49	B	A	A	B	B	A	A	B	A	B	A	A	B	B	B	.	B	B	B	B	B	.	A	B	.
Iapls3-4	B	B	A	A	A	A	A	B	A	B	A	B	A	A	A	.	A	A	A	A	B	.	B	B	.
Pmv4	B	B	A	A	A	A	A	B	A	B	A	B	A	A	A	.	A	A	A	A	B	A	B	B	A
Pmv29	B	A	A	A	A	A	A	B	A	B	A	A	A	.	A	.	A	A	A	A	A	B	B	B	A
D7J2	B	A	A	A	A	A	A	B	A	B	A	B	A	A	B	.	A	A	A	A	A	B	B	B	A
Gpi1	B	A	A	B	A	A	A	B	A	B	A	A	A	A	A	A	A	B	B	B	A	B	B	B	A
Atp1b3	B	B	.	B	A	A	.	B	.	B	A	.	A	A	.	.	.	.	B	B	.	.	B	.	.
Pmv15	B	A	A	A	A	A	A	B	A	B	A	A	A	A	A	.	A	A	A	A	A	B	B	B	A
D7Mit25	B	B	.	A	A	A	.	B	.	B	A	.	A	A	A	.	.	B	B	B	.	.	B	B	A

Strain distribution patterns showing homozygosity for alleles from A/J or C57BL/6 founder in 25 inbred mouse strains comprising the AXB RI set.

Practical Considerations

For those who have cloned a gene and wish to know where it maps, there are several rapid options. First, a number of laboratories have established inter-specific crosses and are interested in mapping as a collaboration. Some of these are listed in Table 1 (p. 10-11) as "Private crosses." The investigator's role is to stick the clone in the mail. For the more independent-minded investigator, there are "Public" resource crosses (Table 1) that provide DNAs from a number of backcross progeny and informatics support in analysing data. The investigator's job is to identify a polymorphism between the mouse strains used in the cross (usually Spret/Ei and C57BL/6 which show a high degree of variation at the sequence level), and to type the approximately 92 DNAs plus control DNAs that comprise the panels. In both cases, these crosses have been previously typed with hundreds of markers, so any new marker typed on the same DNAs will be linked to a previously localized marker.

Recombinant inbred strains provide another alternative for localizing a cloned gene. DNAs from a number of RI strains can be purchased from The Jackson Laboratory. Software for recording and analyzing results and data-bases containing information about hundreds of markers typed previously are available with extensive "Help" files in the MapManager program (see "Table 6" in the Appendix).

Previously established crosses are not useful for mutations that are recognized as a whole animal phenotype (coat color, ataxia, hydrocephalus, behavioral anomalies, blindness, etc.). Mapping this type of marker will require establishment of a new cross. Basic steps in this analysis are listed here; a comprehensive discussion of all possible permutations is beyond the scope of a single lecture. A published source of basic information will be available within the next year in a Cold Spring Harbor Press laboratory manual entitled *Genome analysis: A laboratory manual* edited by B. Birren, E. Green, P. Hieter and R. Myers. Chapter 9, "Genetic and Comparative Mapping in Mice" by Reeves and D'Eustachio provides detailed protocols. The basic steps for mapping a new mutation are as follows:

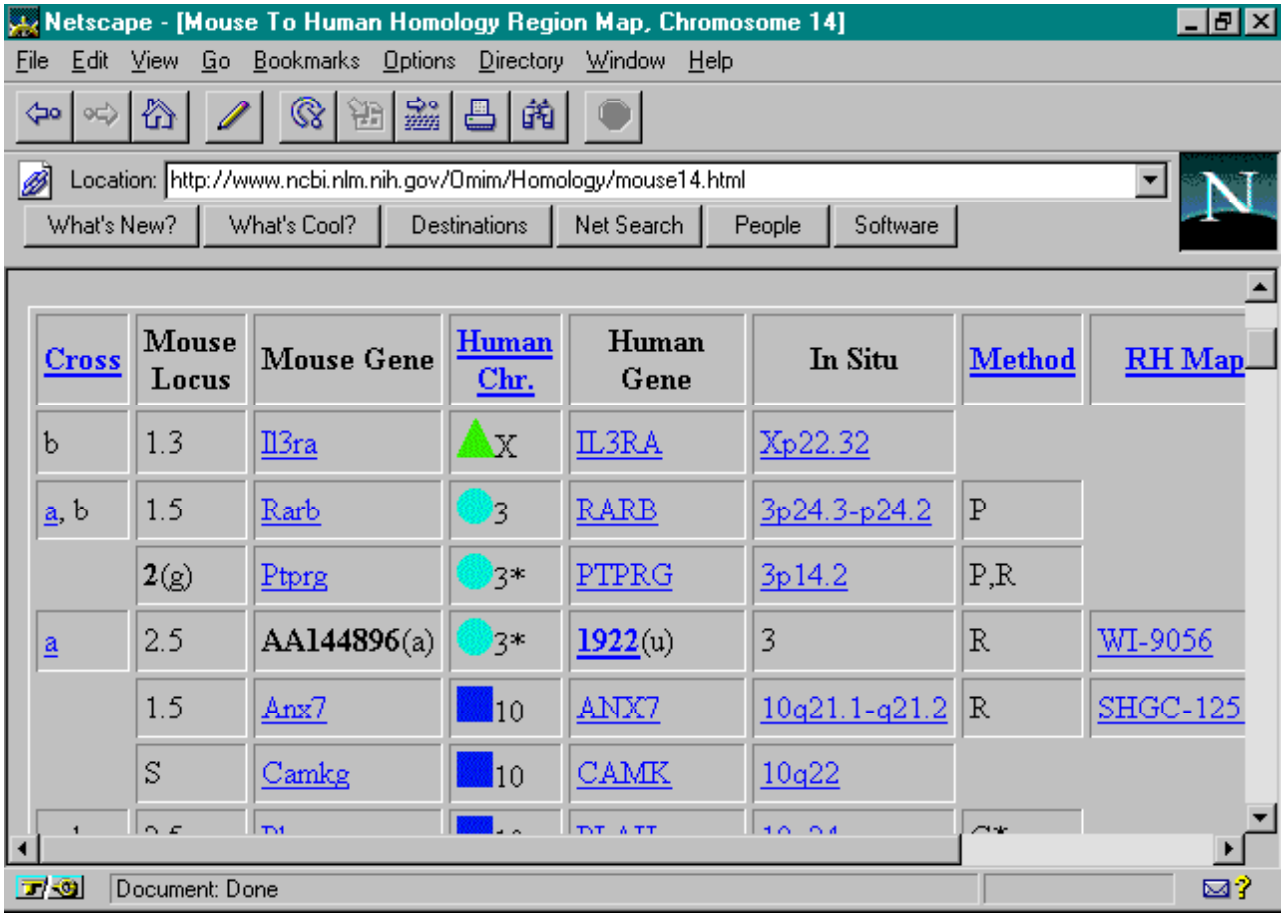
1. Determine mode of inheritance using two generations of test crosses. Possible outcomes:
   - X-linked recessive mutation
   - X-linked dominant or co-dominant mutation
   - Autosomal dominant
   - Autosomal recessive
   - Autosomal co-dominant
   - Recessive lethal mutation

2. Determine an initial localization for the gene using RI strains or by genome scanning of a small backcross or intercross.
   - RI strain analysis is only applicable if an inbred strain expressing the trait has been used in the construction of RI strains.
   - Depending on the mode of inheritance, backcross and intercross strategies will be more or less efficient.

3. Determine fine localization using a large, high resolution cross.

Comparative Mapping

5. Comparative mapping is the prediction of gene location in one species based on the position of that gene in a different species. It derives from the observation that although chromosomes rearrange during evolution, they do not usually scramble. Thus, *orthologous segments* can be identified between species that demonstrate linkage of the same genes, often in the same order. The occurence of a new gene in an orthologous segment implies its localization to the corresponding position in another species.



Cross	Mouse Locus	Mouse Gene	Human Chr.	Human Gene	In Situ	Method	RH Map
b	1.3	Il3ra	X	IL3RA	Xp22.32		
a, b	1.5	Rarb	3	RARB	3p24.3-p24.2	P	
	2(g)	Ptprg	3*	PTPRG	3p14.2	P,R	
a	2.5	AA144896(a)	3*	1922(u)	3	R	WI-9056
	1.5	Anx7	10	ANX7	10q21.1-q21.2	R	SHGC-125
	S	Camkg	10	CAMK	10q22		

Paralogous segments occur within a species. Vertebrate genomes appear to have arisen by a series of whole genome duplications. Remnants of these events are seen as clusters of genes of related structure/function at different sites in the genome. For example, many genes on the short arm of Chr 1 have structurally or functionally related counterparts on the long arm of the chromosome.

Comparison of sequences between species provides important information where a known mutation in one species might be instrumental in defining the gene responsible for a related genetic disease process in another species. These comparisons are also important in identifying the biological roles of normal human genes. For example, isolation of a gene responsible for a mouse mutant in which the immune system is impaired demonstrates the role of the normal homolog of that gene in normal immune system development.

COMPARATIVE MAPPING

Marker type: Any type. It is especially useful for mapping genes identified by a biological function (e.g., mutations). There are several hundred genetically mapped mutations in the mosue affecting virtually every structure, organ system, tissue and developmental stage. The mouse also provides a means of manipulating the germ line to create new mutations at defined positions. Comparative mapping is the mechanism by which biological functions of genes, determined in model systems, can be related to genetic bases of normal and abnormal development and function in humans.

Localization: In terms of mapping, to an orthologous segment. In terms of function, a gene that has been identified as a result of producing a mutation in one species is related to humans via map position and comparative sequence.

Resolution: Highly variable. Currently, comparative maps are used to predict relative positions of genes based on recombinational maps. Some conserved (orthologous) segments are very large, others must be quite small. Comparisons between physical and genetic maps must be made within species as well as between species.

Important sources of comparative map information are available via the Web. Mouse Genome Database, www.informatics.jax.org, contains a wide and deep collection of mouse mapping information of all types. For compataritve mapping questions, a daily update of mouse:human comparisons found at http://www.informatics.jax.org/map.html . Follow links to "Build a graphical map with selected parameters \Linkage." For specific genes, the "Homology" link, http://www.informatics.jax.org/homtools.html, will search for related individual genes in a number of species. NCBI maintains a fast lookup homology table searchable by mouse or human chromosome curated by Dr. Michael Seldin, UC Davis, at http://www.ncbi.nlm.nih.gov/Homology/

The Xref db Cross-Referencing Database is an automated search service comparing predicted proteins from Saccharomyces and other query organisms against all sequences in GenBank (human and many other organisms). Human ESTs with strong similarity to yeast genes are mapped on the Jackson Laboratory Interspecific Backcross (choose "BSS data" at http://lena.jax.org/resources/documents/cmdata/) and on a monochromosomal human somatic cell hybrid panel to predict their locations in the human genome. Since a considerable amount of biological information typically accrues to the yeast genes, this provides annotated data to the human transcription map. Xref db is accessed via world wide web at http://www.ncbi.nlm.nih.gov/XREFdb/ .

Phenoytpe cross-referencing between mouse and human is found in Online Mendelian Inhertitance in Man (http://www.ncbi.nlm.nih.gov/Omim/searchomim.html). This was established by Dr. Victor McKusick as a database of human genetic diseases, and is now extensively cross-referenced to mouse mutations. This database can be searched by a wide variety of terms, including mouse gene symbols, phenotype categories, etc.

Mapping reversed: Creating phenotypes in the mouse.

If one were to select a single genetic tool as the most important for understanding the roles of genes in living organisms, it would likely be analysis of mutations. In a sense, this is the essence of genetics, which is the study of heritable characteristics that vary between individuals. One of the most powerful uses of mouse genetics is the discovery of gene function in complex biological processes. Hundreds of mouse mutations have arisen spontaneously and been "captured" by mouse geneticists. Comparison of these mutant phenotypes, and the underlying genetic bases for them, with normal processes in wild type individuals is a central tenet of genetics.

Increasingly, tools have become available that allow phenotypes to be generated, rather than simply screened for in large breeding colonies. Conventional transgenic mice and gene targeted ("knock out") mice are very familiar examples of genetic engineering in mice. YAC transgenic mice, "transomic" mice carrying large segments of chromosomes or even free-standing chromosomes which allow surveys of selected, large segments of the genome for affects stemming from the aneuploidy represented by these segments.

Targeted chromosome remodeling has been accomplished by targeted insertion of loxP DNA segments which serve as targets for Cre recombinase. When the lox sites are separated by long distances on the same chromosome or on homologous chromosomes, large deletions and insertions can be engineered.

A variety of methods and strategies for mutagenesis originated in large programs established as much as 50 years ago have been revived to create new phenotypes in mouse. The true potential of new chemically-induced mutagenesis strategies coupled with screens for specific mutant loci, configured either as specific locus tests or sensitized screens, are just beginning to show their full potential. Combined with genome remodeling, genomic maps and reagents that greatly facilitate mapping, banks of genes and emerging methods for assessment of gene expression on a whole tissue level, the true potential of mouse as a mammalian model is emerging to a degree that even many mouse geneticists would have found surprising just a few years ago.

Additional Reading:

Lyon, MF, S Rastan, and SDM Brown. 1996. *Genetic Variants and Strains of the Laboratory Mouse*. Third ed., New York: Oxford University Press. From OUP, "This book constitutes a unique source of hard-copy information on mouse variants and is an essential work of reference for biologists who use mouse genetic strains and variants in their research." True.

Reeves, RH and P D'Eustachio. 1998. Genetic and Comparative Mapping in Mouse. In Genome Analysis: A Laboratory Manual, Vol. 4., ed. B Birren, E Green, P Hieter, S Klapholz, R Meyers, H Riethman, and J Roskams. 71-133. Cold Spring Harbor Laboratory Press. Detailed methodology for determining mode of inheritance and mapping Mendelian Traits in the mouse.

Silver, LM. 1995. *Mouse Genetics: Concepts and applications*. NY: Oxford University Press. Well-informed and highly readable discussion about a wide range of applications of mouse genetics.